

Maurer School of Law: Indiana University
Digital Repository @ Maurer Law

Indiana Law Journal

Volume 55 | Issue 3

Article 3

Spring 1980

On the Use of Statistics in Employment Discrimination Cases

Richard M. Cohn

University of Wisconsin- Madison

Follow this and additional works at: <http://www.repository.law.indiana.edu/ilj>



Part of the [Civil Rights and Discrimination Commons](#), and the [Labor and Employment Law Commons](#)

Recommended Citation

Cohn, Richard M. (1980) "On the Use of Statistics in Employment Discrimination Cases," *Indiana Law Journal*: Vol. 55 : Iss. 3 , Article 3.

Available at: <http://www.repository.law.indiana.edu/ilj/vol55/iss3/3>

This Article is brought to you for free and open access by the Law School Journals at Digital Repository @ Maurer Law. It has been accepted for inclusion in Indiana Law Journal by an authorized editor of Digital Repository @ Maurer Law. For more information, please contact wattn@indiana.edu.



JEROME HALL LAW LIBRARY

INDIANA UNIVERSITY
Maurer School of Law
Bloomington

On the Use of Statistics in Employment Discrimination Cases

RICHARD M. COHN*

The use of quantitative information in employment discrimination cases has become increasingly common.¹ Differences in rates of hiring, promotion, demotion or dismissal between minority and majority groups are used to develop a prima facie case of discrimination. Evidence of this type is deemed to represent the outcome of employers' discriminatory practices, including practices the description of which does not indicate any overt discriminatory intent.² This reliance on quantitative information undoubtedly will increase since the new Uniform Guidelines on Employee Selection Procedures (1978) (U.G.E.S.P.) specifically define evidence of discrimination in quantitative terms.³

As a result of this increased emphasis on quantitative information, it is necessary that litigants in discrimination cases, including governmental agencies, become intelligent consumers of such information. Contrary to popular opinion, numbers rarely speak for themselves. Both interpretation and evaluation are required of all but the most elementary quantitative descriptions and analyses. As evidence of this, this article discusses three issues concerning the potential misuse of quantitative information in employment discrimination cases. Each issue illustrates a basic property of quantitative information which affects its probative value. First, the basic issue of when statistical inference is appropriate is considered. To

* A.B. 1970, M.A. 1975, Ph. D. 1977, University of Michigan. Member of the Department of Sociology, and the Center for Demography and Ecology, University of Wisconsin-Madison.

¹ For reviews of the use of quantitative evidence in employment discrimination cases, see Holley & Feild, *Using Statistics in Employment Discrimination Cases*, 4 EMPLOYEE REL. L.J. 43 (1978); Joseph, *Evidence: Statistical Proof in Employment Discrimination Cases*, 28 OKLA. L. REV. 885 (1975); Montlack, *Using Statistical Evidence to Enforce the Laws Against Discrimination*, 22 CLEV. ST. L. REV. 259 (1973).

² As an example of the acceptance of quantitative evidence of discrimination where the description of an employer's hiring policy does not indicate any overt discriminatory practice, see *Lea v. Cone Mills Corp.*, 301 F. Supp. 97 (M.D.N.C. 1969).

³ 43 Fed. Reg. 38,295 (1978) [hereinafter cited as U.G.E.S.P.]. Section 3A of the U.G.E.S.P. states: "Procedure having adverse impact constitutes discrimination unless justified." *Id.* at 38,297. Section 4D defines adverse impact as "[a] selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact." *Id.* Section 16R defines selection rate as "[t]he proportion of applicants or candidates who are hired, promoted, or otherwise selected." *Id.* at 38,308.

illustrate the confusion concerning this issue, a recently published application of an elementary statistical test to employment selection procedures is critically reviewed. The second issue considered is the inappropriate use of particular quantitative measures as evidence of discrimination. Illustrative of this is the new U.G.E.S.P. strategy for determining "prosecutorial discretion" in employment discrimination cases. A simple quantitative model of an employer's hiring rate is developed to indicate the problematic aspects of the U.G.E.S.P. strategy. The final issue discussed is how quantitative analysis may be used to distinguish when a characteristic, such as sex or age, is a "bona fide occupational qualification" from when it is the basis of discrimination.

THE INAPPROPRIATE USE OF STATISTICS AND STATISTICAL INFERENCE

An important misuse of quantitative information in employment discrimination cases may result from the failure to distinguish when such information is a "statistic" as opposed to a "parameter." A statistic is quantitative information which is calculated from sample data. It is used to estimate the corresponding value for the entire population. This value for the population is the parameter. The estimation of the parameter from the statistic is an inferential process. Statistical inference is a probability statement; a statement that a known value of a sample statistic equals the (unmeasured) population parameter with a certain probability. For example, if the results of a sample survey are statistics which indicate women earn less than men, it can be inferred, with a given level of certainty, that in the entire population women, on average, earn less than men. That women actually do earn less than men in the entire population can never be known with complete certainty, *i.e.*, the probability of the truth of the statement is less than 1.0, unless a complete census is carried out.

For a statistic to be an accurate estimate of a parameter, the sample on which the statistic is based must have certain properties. The most important properties are that individuals in the population have a fixed, known and nonzero probability of being included in the sample, and that the sample is representative of the population.⁴ When these properties are absent, the use of a

⁴ Typically these properties of a sample are insured by some form of random selection process. Although sampling procedures can be considerably more complicated, the classic

statistic to infer the value of the parameter is inappropriate.

When using any quantitative information, two basic characteristics of the information must be known. First, it is necessary to know if the information is truly statistical, *i.e.*, calculated from a sample, or if the information is calculated from a complete count (census) of some population. For example, information on male and female promotion rates in a company are parameter values for that company and cannot be viewed as statistics to infer differences in promotion rates for all companies. The specific company's promotion rate information does not constitute a statistic for a sample of companies because the probability of any other company's promotion rates being included as data is zero. Given the information about one company's promotion rates, it is not possible to infer differences in the industry or any other aggregate. The company's promotion rate is a parameter value for that company, known with complete certainty. The information is not a statistic which can be used to estimate the parameter value for any other aggregate with any known certainty.

The second property of any quantitative information which must be known is the definition of the population to which the information applies. In the above example, the population is defined as the employees of the specific company. In situations where the information is truly statistical, the population to which statistical inferences can be made must be precisely defined. For example, if the sample was selected to include only labor force participants, then the population to which inferences can be made is the population of labor force participants, not the total population in a certain geographic area. It is inappropriate to use statistics calculated from this sample to infer parameter values to any population larger or smaller than the population of labor force participants.

An example of the confusion which can result from the failure to determine whether quantitative information is a statistic or a parameter is a recent *Harvard Law Review* article which describes an elementary statistical test purported to be useful in indicating the adverse impact of employment testing.⁵ In the article, the author,

example of drawing names from a hat to produce a sample is instructive. Having each individual's name appear on only one slip insures that each member of the population will have a fixed, known, and nonzero probability of being selected (in this case each member has an equal chance of selection). Thoroughly stirring the names and selecting them blindly insures that the sample will be representative of the population. Details on the necessary properties of a sample and descriptions of sampling techniques can be found in most elementary statistics texts and research methods texts. See, *e.g.*, L. KISH, *SURVEY SAMPLING* (1965).

⁵ Shoben, *Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under*

Professor Elaine Shoben, proposes that the difference in passing rates on an employment selection test between minority and majority applicants be considered a statistic and used to estimate whether the test has an adverse impact in the population.⁶ This proposed use of the difference in passing rates information for a specific employer as a statistic is contrasted with the U.G.E.S.P.'s "4/5ths rule" to indicate adverse impact of the employment test. With the guideline rule, a difference of twenty percent or more in the proportion passing the test between minority and majority applicants is viewed as indicating adverse impact.⁷

The Shoben article argues that the proposed statistical test is superior to the use of the "4/5ths rule" which is not liable to statistical inference. While the U.G.E.S.P. indicate that the "4/5ths rule" is to be used in lieu of statistical tests, the guidelines are vague as to when, if ever, the "4/5ths rule" is to be applied. As noted in section 4D of the U.G.E.S.P., a statistically significant difference in passing rates smaller than 0.2 may be used as evidence of adverse impact when there is a large number of applicants. In cases where there is a small number of applicants, a difference in passing rates greater than 0.2 may not be used to indicate adverse impact if the difference is not statistically significant.⁸ Thus, the number of applicants an employer processes may determine whether or not adverse impact of the selection procedures is indicated, regardless of the true discriminatory nature of the test, be-

Title VII, 91 HARV. L. REV. 793 (1978). Adverse impact, as used in the Shoben article and defined in § 16B of the U.G.E.S.P., is "[a] substantially different rate of selection in hiring, promotion, or other employment decision which works to the disadvantage of members of a race, sex, or ethnic group." 43 Fed. Reg. 38,307 (1978).

⁶ The Shoben article suggests the use of the difference in proportions test, where the observed difference in proportions passing the test between minority and majority applicants of an employer is used to estimate whether for the entire population there exists a difference in passing rates. For a description of the test, see Shoben, *supra* note 5, at 797-805, or any elementary statistics text.

⁷ See note 3 *supra*.

⁸ Section 4D of the U.G.E.S.P. states:

Smaller differences [than 0.2] in selection rate may nevertheless constitute adverse impact, where they are significant in both statistical and practical terms or where the user's actions have discouraged applicants disproportionately on grounds of race, sex, or ethnic group. Greater differences in selection rate may not constitute adverse impact where the differences are based on small numbers and are not statistically significant . . .

43 Fed. Reg. 38,297 (1978). See also Questions and Answers 22 & 24, interpreting and clarifying the U.G.E.S.P. 44 Fed. Reg. 11,999 (1979). As the federal enforcement agency has discretion in choosing whether a test of statistical significance or the "4/5ths rule" is used to indicate adverse impact, considerable uncertainty among employers as to when they are in violation of the U.G.E.S.P. can be expected.

cause alternate decision rules may be applied to the quantitative information.

Any statistical test, the one proposed in the Shoben article or the test proposed as an alternative to the "4/5ths rule" in the U.G.E.S.P., is inappropriate to indicate adverse impact. The employer's quantitative information on passing rates is not a statistic, but rather a population parameter. As such, the use of any test of statistical significance of the differences in passing rates to infer to a population larger than that of the individuals who took the test is inappropriate. In the Shoben article it is argued that the test-takers can be considered a sample of all potential job applicants in the employer's "labor pool," *i.e.*, they are a sample of a population defined as the potential job applicants of the employer. In fact, the applicants who took the test are not a sample of anything, for there is no fixed or known probability by which all potential applicants become test-takers, *i.e.*, become part of the sample. Those individuals who took the test define the population of test-takers. No statistical inference relating a difference in their observed passing rates to an unmeasured difference in passing rates in a larger population is warranted.⁹

A similar argument applies to the U.G.E.S.P.'s suggested use of statistical tests as an alternative to the "4/5ths rule" in determining adverse impact. The information of the employer's passing rates is not sample information, it is a population parameter.¹⁰ There is no appropriate use of a test of statistical significance with this information because there is no larger population to which to infer. Regardless of the number of applicants taking the test, the difference in passing rates between minority and majority applicants has the same value in indicating adverse impact.¹¹ As an al-

⁹ Shoben, *supra* note 5, at 801, notes that as a "sample," the actual test-takers may not be representative of the "population" of all potential test-takers. This is likely true, but it is irrelevant given that the test-takers do not constitute a sample in the first place.

¹⁰ The definition of the population need not be limited to only those applicants taking the test at a single administration. The population may be defined as all the applicants taking the test during a specific year. For purposes of evaluating adverse impact of the test, a change in the contents of the test would require a new population of applicants be defined.

¹¹ With statistical information, the certainty with which one can infer that the sample statistic equals the population parameter increases with the size of the sample. Thus, when quantitative information is truly statistical, its acceptance by the court or governmental enforcement agency may be questioned if the sample size is deemed inadequate. Quantitative information which is calculated from an entire population cannot be disputed on the basis of the number of observations. Thus, in the above example of the difference in passing rates to indicate adverse impact, because the information is a population parameter, the number of observations (test-takers) is irrelevant to the question of adverse impact. The

ternative to the "4/5ths rule," the use of a test of the statistical significance of the difference in passing rates is inappropriate.¹²

The level of statistical significance indicates the certainty with which a difference in passing rates observed in a sample can be inferred to exist in a population.¹³ Statistical significance is meaningless when not referring to the relation of a sample value (statistic) to a population value (parameter). The failure to note that the information obtained from an employer is not sample information can produce erroneous uses of quantitative information such as found in the Shoben article and section 4D of the U.G.E.S.P.

The determination of whether the information is derived from a sample is relevant to the quality of the evidence indicating discrimination. If the information on the test-takers' performance is erroneously conceptualized as being sample data from a population of all potential test-takers, the test of statistical significance addresses the question: "Is it likely there would be such a difference in observed test scores if the test has no adverse impact in the entire population?" Not only are the observed scores inadequate evidence to answer this question when they are viewed as sample results, due to the unknown bias of the "sample," but this question is not the legal question at issue. The question at issue is whether

census of Liechtenstein is as informative as the census of China.

However, there is a reason for requiring a minimum number of test-takers in the population which is unrelated to the issue of statistical significance. Since several factors, other than race, may determine test performance, in comparing the passing rates of minority and majority applicants, it is important to control for racial differences in these nonracial factors, *e.g.*, education, experience. If there are racial differences in these factors, there should be an adequate number of observations of minority and majority group applicants with similar qualifications to allow for racial comparisons when these qualifications are held constant. The importance of controlling for racial differences in job qualifications when evaluating racial differences in hiring is discussed later in this article. See notes 25-38 & accompanying text *infra*.

¹² Rather than not applying the "4/5ths rule" because it is too "stringent" for employers with few applicants and too "lenient" for employers with many applicants, an alternative form of the rule might be used. If federal enforcement agencies consider it appropriate to include the number of applicants, as well as the difference in selection rates, in determining adverse impact, a graduated "rule" of the difference in selection rates is appropriate. The graduated structure of the rule could be progressive, similar to the structure of the income tax rate. Thus, the larger the number of applicants processed by an employer, the smaller the difference in selection rates necessary to indicate adverse impact.

¹³ Actually, the common manner by which the level of statistical significance is stated denotes the degree of certainty of no difference not existing in the population. Thus, the 0.05 level of significance indicates a certainty of 1 out of 20 that no difference in selection rates does not exist in the population. The reason for the double negative is that statistical tests are formulated as a rejection of a (null) hypothesis which is counter to the actual hypothesis of interest.

or not the company's test has produced a differential passing rate between groups. This question is nonprobabilistic, has no reference to test-takers at any other point in time, and is unequivocally answered by treating the quantitative information as population data.

DEFINING DISCRIMINATION WITH AN INAPPROPRIATE MEASURE

The second misuse of quantitative information in employment discrimination cases is the inappropriate choice of a particular measure as evidence of discrimination. Various ratios have been used to indicate disproportionate hiring or employment of minority group members. Perhaps the most commonly used parameter is the ratio of the proportion of hires (H) in the population or labor pool (L) between minority and majority groups, *i.e.*, $[(H/L)_{\text{blacks}} / (H/L)_{\text{whites}}]$. A value of this ratio of less than one has been used as evidence of discrimination.¹⁴

The use of this ratio as evidence of an employer's discriminatory practices can be misleading. The proportion hired for a particular group can be decomposed as the product of several proportions:

$$(H/L) = (H/S)(S/Q)(Q/A)(A/L)$$

where:

L=number of group members in the labor pool

A=number of applicants

Q=number of applicants who are qualified for the job

S=number of qualified applicants who pass the employer's selection procedures

H=number who are hired.

Ratios of each of these proportions for minority and majority groups can be indicated as

$$[(H/L)_{\text{blacks}} / (H/L)_{\text{whites}}] = [(H/S)_{\text{blacks}} / (H/S)_{\text{whites}}] [(S/Q)_{\text{blacks}} / (S/Q)_{\text{whites}}] [(Q/A)_{\text{blacks}} / (Q/A)_{\text{whites}}] [(A/L)_{\text{blacks}} / (A/L)_{\text{whites}}]$$

¹⁴ See, *e.g.*, Boston Chapter, NAACP v. Beecher, 504 F.2d 1017 (1st Cir. 1974); Ochoa v. Monsanto Co., 473 F.2d 318 (5th Cir. 1973); Castro v. Beecher, 459 F.2d 725 (1st Cir. 1972); Parham v. Southwestern Bell Tel. Co., 433 F.2d 421 (8th Cir. 1970). In the above cases, the total population, rather than the size of the relevant labor pool, was used to indicate L. The inappropriateness of using the total population instead of a measure of the size of the labor pool is noted in Hazelwood School-Dist. v. United States, 433 U.S. 299, 308 n.13, 310-12 (1977).

To be consistent with the form of the measure used in court cases and the U.G.E.S.P., H represents the number of individuals who are hired by the company. A more precise indicator of an employer's discriminatory practices would be obtained if H stood for the number of individuals who are offered employment, regardless of whether they accept the offer.

L)_{whites}].

The ratio of the proportions (A/L) indicates the difference between minority and majority groups in their application for the job. The ratio of the (A/L) proportions can be assumed to be only partially influenced by the employer's recruitment efforts. A series of other factors, including alternative job opportunities in the labor market, may differentially determine the proportion (A/L), and consequently (H/L), for minority and majority groups. Thus the use of the ratio of the proportions (H/L) to indicate the employer's discriminatory practices, without also indicating the proportions (A/L), can be misleading.¹⁵

Perhaps recognizing the fallibility of this measure, the U.G.E.S.P. indicate that the appropriate indicator of discriminatory practice is not the ratio of the hiring proportions (H/L), but the ratio of the "selection rates."¹⁶ Using the notation described above, the selection rate may be indicated as (H/A) and decomposed as $(H/A) = (H/S)(S/Q)(Q/A)$.¹⁷

The rate at which minority applicants are selected, *i.e.*, (H/A)_{blacks}, is compared to the rate for majority group applicants, (H/A)_{whites}. If the ratio of these rates is less than 0.8 the "4/5ths rule" is violated and adverse impact of the employer's selection procedures is indicated. While the selection rate (H/A) is a more accurate measure of the employer's discriminatory practices than the hiring rate (H/L), it too can be misleading. To show this, the determinants of the selection rate's components must be considered.

The proportion of applicants truly qualified for the job, (Q/A), is generally unknown to the employer.¹⁸ The employer has little or no control over this proportion. There is no reason to assume that the proportion (Q/A) is the same for minority and majority groups. Due to unequal opportunities for schooling, job training programs, accumulation of relevant job experience, and other factors, it is possible that the minority group has a lower (Q/A) proportion than

¹⁵ The misleading nature of using the ratio of proportions (H/L) without also reporting the ratio of the proportions (A/L) has been noted by the courts. See *Richardson v. Indiana Bell Tel. Co.*, 2 Fair Empl. Prac. Cas. (BNA) 797 (S.D. Ind. 1970).

¹⁶ See note 3 *supra*, for the U.G.E.S.P. definition of "selection rate."

¹⁷ For definitions of what constitutes an "applicant" and a "candidate," see Question and Answer 15, interpreting the U.G.E.S.P. 44 Fed. Reg. 11,998 (1979).

¹⁸ Unless an analysis of the job produces a definitive list of the qualifications for the particular job and these qualifications are characteristics of applicants which are vested in the applicants, rather than acquired after employment, the rate (Q/A) can never be known with certainty.

the majority group. Conversely, it is equally possible that because of other employers refusing to hire qualified minority workers, there is a surplus of qualified minority applicants for a specific employer and therefore the (Q/A) proportion of the minority group is greater than the (Q/A) proportion of the majority group.

The proportion of qualified applicants who pass the employer's selection procedures, (S/Q), can be assumed to be equal for minority and majority groups if the procedures are fairly administered and actually measure the qualifications for the job.¹⁹ Since the number of qualified applicants, Q, is unknown, the ratio of the (S/Q) proportions of the groups can never be calculated directly. Yet it is a reasonable assumption, given fair and relevant selection procedures, that minority and majority applicants of equal qualifications will pass the selection procedures at the same rate.

The proportion of those successfully completing the selection procedures who are hired, (H/S), is completely determined by the employer. This proportion is a known, directly observable quantity. Any difference in the proportion between minority and majority groups, *i.e.*, a ratio of proportions not equal to one, can be viewed as evidence of discrimination.²⁰

In evaluating the adverse impact of an employer's selection procedures and determining "prosecutorial discretion," the U.G.E.S.P. indicate a "bottom line" strategy will be used.²¹ This strategy con-

¹⁹ A perfectly accurate selection procedure would completely distinguish qualified from unqualified applicants and thus (S/Q) would be equal to 1.0. The requirements for validation of selection procedures described in the U.G.E.S.P. do not require such a stringent level of validation. As indicated in § 5B of the U.G.E.S.P., "[e]vidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance." 43 Fed. Reg. 38,298 (1978). Thus, the proportion (S/Q) may be considerably lower than 1.0. Part of the requirement for selection procedure validation includes producing evidence of "fairness," *i.e.*, the selection procedure does not differentially indicate qualification for the job between minority and majority applicants. See § 14B8 U.G.E.S.P., 43 Fed. Reg. 38,299 (1978). This part of the validation process is to insure that (S/Q)_{blacks} equals (S/Q)_{whites}.

²⁰ A ratio of the proportions (H/S) greater than 1.0, *i.e.*, [(H/S)_{blacks}/(H/S)_{whites}] > 1.0, indicates "affirmative action" on the part of the employer.

²¹ Section 4C of the U.G.E.S.P. states, in part,

The 'bottom line.' If the information called for by sections 4A and B above shows that the total selection process for a job has an adverse impact, the individual components of the selection process should be evaluated for adverse impact. If this information shows that the total selection process does not have an adverse impact, the Federal enforcement agencies, in the exercise of their administrative and prosecutorial discretion, in usual circumstances, will not expect a user to evaluate the individual components for adverse impact, or to validate such individual components, and will not take enforcement action

sists of examining the ratio of the selection rates for minority and majority groups $[(H/A)_{\text{blacks}}/(H/A)_{\text{whites}}]$. If this ratio is greater than 0.8, *i.e.*, if the "4/5ths rule" is not violated, the U.G.E.S.P. indicate that the individual components of the selection procedure will not be evaluated for adverse impact and enforcement action is not likely.²²

The use of this "bottom line" strategy can lead either to the false charge of adverse impact or to the conclusion that no adverse impact exists when, in fact, the employer's selection procedure is discriminatory. These results are possible because of the misuse of the quantitative information of selection rates (H/A) to indicate adverse impact. This problematic feature of the "bottom line" strategy may be illustrated by two examples. The cases of two hypothetical employers are described in Table I.

based upon adverse impact of any component of that process, including the separate parts of a multipart selection procedure or any separate procedure that is used as an alternative method of selection.

43 Fed. Reg. 38,297 (1978).

²² Exceptions to the use of the "bottom line" strategy are noted in § 4C of the U.G.E.S.P. The relevant part of the section states:

However, in the following circumstances the Federal enforcement agencies will expect a user to evaluate the individual components for adverse impact and may, where appropriate, take enforcement action with respect to the individual components: (1) where the selection procedure is a significant factor in the continuation of patterns of assignments of incumbent employees caused by prior discriminatory employment practices, (2) where the weight of court decisions or administrative interpretations hold that a specific procedure (such as height or weight requirements or no-arrest records) is not job related in the same or similar circumstances. In unusual circumstances, other than those listed in (1) and (2) above, the Federal enforcement agencies may request a user to evaluate the individual components for adverse impact and may, where appropriate, take enforcement action with respect to the individual component.

Id. Neither of these exceptions is relevant to the critique of the "bottom line" strategy described herein.

TABLE I
EXAMPLES OF ERRONEOUS EVALUATIONS OF ADVERSE IMPACT
USING THE U.G.E.S.P. "BOTTOM LINE" STRATEGY

| | Case 1: False Charge of Adverse Impact | | | Case 2: Failure to Discover Adverse Impact | | |
|-----|---|--------|--------------------------|---|--------|--------------------------|
| | Whites | Blacks | Ratio (Blacks/Whites) | Whites | Blacks | Ratio (Blacks/Whites) |
| A | 1000 | 200 | | 1000 | 200 | |
| Q | 900 | 120 | | 600 | 180 | |
| S | 630 | 84 | | 600 | 180 | |
| H | 567 | 84 | | 600 | 108 | |
| Q/A | 0.90 | 0.60 | 0.67 | 0.60 | 0.90 | 1.50 |
| S/Q | 0.70 | 0.70 | 1.00 | 1.00 | 1.00 | 1.00 |
| H/S | 0.90 | 1.00 | 1.11 | 1.00 | 0.60 | 0.60 |
| H/A | 0.57 | 0.42 | 0.74 | 0.60 | 0.54 | 0.90 |
| H/Q | 0.63 | 0.70 | 1.11 | 1.00 | 0.60 | 0.60 |

where:

A = number of applicants

Q = number of applicants qualified for the job

S = number of applicants who successfully pass employer's selection procedures

H = number of applicants who are hired

H/A = "selection rate"

In Case 1, the employer would be falsely charged with an adverse impact of his selection procedures because the ratio of the selection rates (reported in the third column of the Table) is less than 0.8. Yet qualified blacks are passing the employer's selection procedures at the same rate as qualified white applicants and the employer is hiring selected black applicants at a rate greater than selected white applicants, *i.e.*, the employer has a program of affirmative action. The false charge of adverse impact of the selection procedures results from the proportion of black applicants who are qualified for the job being only two-thirds that of the proportion of white applicants.

In litigation of the charge of adverse impact, the employer in Case 1 could point to the following facts to support his case:

1. the affirmative action evident by the ratio $[(H/S)_{\text{blacks}} / (H/S)_{\text{whites}}]$ equaling 1.11;

2. qualitative evidence of fair administration of the selection procedures to infer $(S/Q)_{\text{blacks}} \text{ equals } (S/Q)_{\text{whites}}$.

Whether the qualitative evidence of the second fact can be successful in countering the quantitative evidence that $[(S/A)_{\text{blacks}}/(S/A)_{\text{whites}}] \text{ equals } 0.67$ (a violation of the "4/5ths rule") is not discussed in the U.G.E.S.P.

In Case 2, the employer is not charged with adverse impact under the "bottom line" strategy because the ratio of selection rates is greater than 0.8. Yet the employer is hiring blacks who have successfully passed the employment selection procedures at a rate only sixty percent that of whites. The failure of the "bottom line" strategy to discover this discriminatory practice is a result of the proportion of black applicants who are qualified for the job being fifty percent greater than the proportion of white applicants.²³

The basic flaw of the U.G.E.S.P.'s "bottom line" strategy is that it uses too highly aggregated quantitative information to indicate discrimination. The strategy's reliance on the ratio of the selection rates (H/A) to indicate an employer's discriminatory practices is based on the assumption that the ratio of the (Q/A) proportions is 1.0. This is always an untested and questionable assumption. The most accurate single measure of the discriminatory nature of the employer's hiring process, the ratio of the proportion of qualified applicants who are hired, (H/Q) , is not directly observable. Thus, to accurately detect and evaluate the adverse impact of the employer's selection process requires the use of both disaggregated measures of the selection process, *e.g.*, (H/S) , and direct qualitative analyses of the procedures used in the process.²⁴

ANALYTICAL PROCEDURES TO DETECT DISCRIMINATION

The third issue in the use of quantitative information in employment discrimination cases concerns how the information is properly analyzed to indicate discrimination. When quantitative information indicates differential treatment by an employer of

²³ The exceptions to the use of the "bottom line" strategy, *see* note 22 *supra*, would not make the employer in Case 2 subject to evaluation of the components of his selection process. Thus the employer's discriminatory practice of hiring blacks who have passed the selection procedures at a lower rate than whites would go undetected.

²⁴ To simplify the illustration, the hiring model described above includes only one term to indicate selection testing (S/Q) . The model can easily be expanded to include separate terms for each stage of the selection process, *e.g.*, interviews, paper and pencil tests.

individuals who differ only with respect to race, it is a reasonable assumption that discrimination has occurred. The basis of this assumption is that an individual's race has no effect on productivity. Therefore, the differential treatment by race cannot be motivated by an acceptable economic reason.²⁵ It is so implausible that race can determine a worker's productivity that race is not liable to be a "bona fide occupational qualification" (BFOQ).²⁶ A similar categorical determination of irrelevance to productivity cannot be made for the characteristics of sex and age. Both sex and age can be assumed to influence productivity in certain instances. As a result, the equal employment opportunity guidelines on sex and age discrimination include discussions indicating when these characteristics constitute a BFOQ.²⁷ Since age and sex, but not race, can influence productivity, differential treatment by sex and age may have an economically rational basis. This produces a difference in the type of quantitative information which is appropriate to indicate sex and age discrimination from the type of evidence appropriate to indicate race discrimination.

The basic distinction in types of evidence involves the analytical concept of an "intervening variable." An intervening variable is a variable which is the source of a relationship between two other variables. As an illustration, consider a relationship between race and productivity. In this hypothetical example, it is found that blacks, on average, have lower productivity than whites. It is also found that the level of education of a worker is positively related to productivity and that the average level of education is lower for the group of blacks. If the relationship between race and productivity is examined separately for each level of education, it is found that there is no difference in productivity between blacks and whites.²⁸ The productivity difference between blacks and whites

²⁵ Defining economic discrimination as existing when factors, other than productivity, determine a worker's pay or the probabilities of hiring, promotion or dismissal is common in the social scientific literature on discrimination. See Aigner & Cain, *Statistical Theories of Discrimination in Labor Markets*, 30 INDUS. & LABOR REL. REV. 175 (1977).

²⁶ The differential treatment of a minority group as a result of a "business necessity" does not contradict this. See *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). The concept of a business necessity to hire a disproportionate number of whites results from an insufficient number of black applicants having the requisite qualifications for the particular job.

²⁷ See Equal Employment Opportunity Commission, 29 C.F.R. § 1604.2 (1979), for the guidelines on sex as a BFOQ. See EEOC, Interpretations, 29 C.F.R. § 860.102 (1979), for discussion of age as a BFOQ relating to the Age Discrimination in Employment Act of 1967.

²⁸ This technique of examining a relationship between two variables while holding constant the value of a third variable is called "statistical control." Discussions of statistical control can be found in most elementary statistical texts.

has been explained by the racial difference in educational attainment. Education is the intervening variable in the relationship between race and productivity.

For a characteristic such as sex or age to be a BFOQ, there must be a strong relationship between the characteristic and the job qualification which is the intervening variable in the relationship between the characteristic and productivity.²⁹ The strength of the relationship between the characteristic and the intervening variable can indicate the degree to which the use of the characteristic as a BFOQ is valid.

In certain instances, sex and age can be considered determinants of productivity because there is a strong relationship between sex or age and a qualification for the job.³⁰ In such instances it is necessary to evaluate whether differential treatment in wage, hiring, promotion or dismissal rates is truly a result of the relationship between the characteristic and the job qualification, or if it is a result of discrimination. Quantitative analyses can be useful in making this distinction.

As an example of this type of analysis, consider a hypothetical

²⁹ As an example, consider the relationships between an actor's age, athletic ability and the quality of a performance as Tarzan of the Jungle. The quality of the performance is negatively related to an actor's age. This is the result of the reasonably strong relationship between age and the intervening variable, athletic ability. As a result of the relationship between the true occupational qualification of athletic ability and the characteristic of age, youth is considered a BFOQ to portray Tarzan. See EEOC, Interpretations, 29 C.F.R. § 860.102e (1979).

³⁰ If there is a "perfect" relationship between the characteristic and the job qualification, then the characteristic itself is a valid BFOQ. For example, the qualification for modeling women's fashions is a female anatomy, therefore, being female is a BFOQ. At the other extreme there are job qualifications which have no relationship to sex, *e.g.*, cognitive ability. Between these extreme examples is a variety of job qualifications which exhibit some relationship to sex, *e.g.*, physical strength. There is no standard ruling by the courts as to how strong a relationship between a characteristic and a job qualification there must be before the characteristic is defined as a BFOQ. In *Rosenfeld v. Southern Pac. Co.*, 444 F.2d 1219 (9th Cir. 1971), that court held that "sexual characteristics, rather than characteristics that might . . . correlate with a particular sex, must be the basis of the BFOQ exception." *Id.* at 1225. There the court indicated that only when there is a perfect relationship between the characteristic (sex) and the job qualification is a BFOQ acceptable. Compare this ruling with the ruling in *Massachusetts Bd. of Retirement v. Murgia*, 427 U.S. 307 (1976). In that case the Supreme Court accepted the characteristic of age as a BFOQ in lieu of the actual job qualification of physical fitness. The Court stated:

That the State chooses not to determine fitness more precisely through individualized testing after age 50 is not to say that the objective of assuring physical fitness is not rationally furthered by a maximum age qualification.

Id. at 316. Thus in *Murgia*, a less than perfect relationship between a characteristic (age) and a job qualification (physical fitness) was accepted as a basis of making youth (being less than 50 years old) a BFOQ.

employer who is hiring for a particular job in the company. An analysis of the job indicates that there are three qualifications necessary for success in the position. The employee must (1) be at least a high school graduate, (2) be willing to stay on the job for the next five years, and (3) be able to lift a 100-pound weight. The employer decides that each of these qualifications should be weighted equally in determining expected productivity, and therefore the probability of hiring an applicant. The simple decision-making model used by the employer to determine who should be hired can be described by the following formula:

$$\text{HIRE} = (\text{GRAD}/3) + (\text{STAY}/3) + (\text{LIFT}/3)$$

where:

HIRE=the probability that the applicant is hired³¹

GRAD=1, if the applicant has at least a high school education
0, if the applicant has less than a high school education

STAY=the probability that the applicant is willing to stay on the job for the next five years. This probability is calculated as $[1-(\text{applicant's age}/70)]$

LIFT=1, if the applicant is able to lift a 100-pound weight
0, if the applicant is not able to lift a 100-pound weight.

The employer assumes that women, on the average, cannot pass the weightlifting test. The employer also knows that the probability of a new employee staying on the job for the next five years is negatively related to the employee's age, as described by the formula: $\text{STAY}=[1-(\text{applicant's age}/70)]$. Given this knowledge, the employer wants to determine whether being male and young are BFOQs.

To resolve this question empirically, the employer analyzes the data available from his pool of applicants for the job. A description of the applicant pool is presented in Table II. Each applicant's sex, age, educational attainment, whether the applicant passed the weightlifting test, and the probability of staying on the job for five years is reported in Table II. In addition, the probability of hiring

³¹ Viewing the decision of hiring a particular applicant as a probability, rather than a dichotomous choice of hire or not hire, has certain heuristic value. However, the decision-making model described in the example may be altered to indicate whether or not an applicant is hired without affecting the conclusions derived from the example. Alternatively, the probability of hiring (HIRE) can simply be thought of as representing the rank ordering of applicants for the position; with the greater the value of HIRE, the higher the rank of the applicant.

the applicant (HIRE), calculated using the decisionmaking model described above, is reported.

TABLE II
APPLICANT POOL

| Applicant | Sex | Age | Job Qualifications | | | Hiring Probability (HIRE) |
|-----------|-----|-----|-------------------------|-------------------------------|--|---------------------------------|
| | | | H.S. Grad. (GRAD) | Passed Lift Test (LIFT) | Probability of Staying 5 Years (STAY) | |
| 1 | M | 35 | YES | YES | .50 | .83 |
| 2 | M | 35 | YES | NO | .50 | .50 |
| 3 | M | 40 | NO | YES | .43 | .48 |
| 4 | F | 40 | YES | NO | .43 | .48 |
| 5 | F | 40 | YES | YES | .43 | .81 |
| 6 | M | 45 | NO | YES | .36 | .45 |
| 7 | F | 45 | NO | NO | .36 | .12 |
| 8 | F | 50 | NO | NO | .29 | .10 |
| 9 | F | 55 | NO | NO | .21 | .07 |
| 10 | M | 60 | NO | YES | .14 | .38 |
| 11 | F | 60 | NO | NO | .14 | .05 |
| 12 | M | 65 | NO | YES | .07 | .36 |
| 13 | F | 65 | YES | NO | .07 | .36 |
| 14 | F | 65 | YES | YES | .07 | .69 |
| 15 | M | 70 | YES | NO | .00 | .33 |
| 16 | M | 70 | YES | YES | .00 | .67 |

Using these data, the employer calculates the average probability of hiring a man or a woman. The average probability of hiring a male applicant is 0.5; the average probability of hiring a female applicant is 0.33. Since the probability of hiring is based on the expected productivity in the job, this sex difference in the hiring probability represents an expected sex difference in productivity.

To determine whether this sex difference is adequate grounds for making being male a BFOQ, the employer needs to determine the cause of the sex difference in the hiring probability (expected productivity). To do this, the employer analyzes the data for relationships between sex and the three job qualifications (GRAD, STAY, and LIFT). An appropriate technique for this is correlational analysis.³² The correlation between sex and educational at-

³² A correlation coefficient indicates the strength of the relationship between two variables. The coefficient has a range of values from -1.0 to + 1.0. A value of -1.0 indicates a perfect negative relationship between two variables; a coefficient equal to + 1.0 indicates a perfect positive relationship. A correlation coefficient equal to 0.0 indicates the complete absence of a relationship between two variables. A description of how correlation coefficients are calculated can be found in most statistics texts.

tainment (GRAD) and the correlation between sex and the probability of staying on the job for five years (STAY) are both 0.0, indicating there is no sex difference in either of these qualifications.³³ The correlation between sex and passing the weightlifting test (LIFT) is 0.5, indicating a moderate sex difference in this qualification.³⁴ The ability to pass the weightlifting test is the intervening variable between sex and the hiring probability which causes the sex difference in that probability. This can be formally shown by comparing the correlation coefficient between sex and the hiring probability (HIRE) with the correlation between sex and the hiring probability when differences in the ability to pass the weightlifting test are controlled for. This correlation between sex and the hiring probability after taking into account the weightlifting qualification is called a partial correlation.³⁵ The correlation between sex and the hiring probability is 0.34, indicating that female applicants are less likely to be hired than male applicants. After controlling for whether or not an applicant passed the weightlifting test, the partial correlation between sex and the hiring probability is 0.0. This indicates that the sex difference in the hiring probability is completely due to the sex difference in the ability to pass the weightlifting test.

The moderate strength of the relationship between sex and the job qualification of passing the weightlifting test (the correlation coefficient of 0.5) would make it difficult for the employer to substantiate a claim that being male is a BFOQ for the job. Women should not be categorically excluded from the applicant pool.

A similar analysis is appropriate to determine whether youth is a BFOQ. Using the same data for the applicant pool, there is a correlation of -0.24 between the applicant's age and the probability of being hired. This indicates that the older an applicant is, the lower

³³ The lack of relationships between sex, educational attainment and the probability of staying on the job for five years was built into the data to simplify the example. Sex differences in these qualifications and relationships among the qualifications could exist in data for an actual applicant pool. If such relationships do exist, the basic analytical technique used in the example is still appropriate, although the calculations become considerably more complex.

³⁴ Another method of noting the sex difference in the weightlifting job qualification is to calculate the percentages of male and female applicants who passed the test. Seventy-five percent of the male applicants passed the test compared to only twenty-five percent of the female applicants.

³⁵ The value of a partial correlation coefficient has the same interpretation as the value of a simple correlation coefficient, see note 32 *supra*, i.e., its range of values is from -1.0 to +1.0, with 0.0 indicating the lack of a relationship after controlling for the intervening variable.

is the probability of being hired. An analysis of the correlations between age and the three job qualifications reveals correlation coefficients of 0.0 between age and educational attainment (GRAD) and between age and whether the applicant passed the weightlifting test (LIFT). There is a correlation coefficient of -1.0 between age and the probability that the applicant will stay on the job for five years.³⁶

In this extreme example, youth completely determines this job qualification. That the age difference in the probability of being hired results from age determining this job qualification can be demonstrated by the partial correlation between age and the hiring probability controlling for the probability of staying on the job for five years (STAY). This partial correlation equals 0.0, indicating that the simple correlation between age and the probability of being hired (-0.24) is a result of the intervening variable STAY.

Therefore, youth is a BFOQ because it completely determines one of the job qualifications. Given that staying on the job for five years is truly a determinant of productivity and the probability of staying is accurately described by the formula used to calculate STAY, the employer might properly use youth as a BFOQ.³⁷

The results from this type of analysis can be used as quantitative evidence of the validity or invalidity of the use of a particular characteristic as a BFOQ. An additional use of this analysis is to provide direct quantitative evidence of discrimination in situations where an applicant's characteristic is related to a qualification of the job and is the object of an employer's discrimination.

³⁶ As in the example of sex differences, the lack of a relationship between age and both educational attainment and whether the applicant passed the weightlifting test is an artificial construction used to simplify the example. The perfect relationship between age and the probability of staying on the job for five years is also artificially built into the data. Its purpose is to indicate an extreme example of when a characteristic of an applicant and a job qualification are perfectly related. It is this type of perfect relationship that was considered necessary for a characteristic to be a BFOQ in *Rosenfeld v. Southern Pac. Co.*, 444 F.2d 1219 (9th Cir. 1971).

³⁷ With youth as a BFOQ, the employer's decisionmaking model may be described by the formula:

$$\text{HIRE}\dagger = (\text{GRAD}/3) + ((70 - \text{AGE})/210) + (\text{LIFT}/3)$$

where AGE is the applicant's age. In this example, HIRE†, the probability an applicant is hired when youth is a BFOQ, equals HIRE as reported in Table II. This is because STAY is completely determined by the applicant's age. It is worth noting that using youth as a BFOQ does not categorically exclude an applicant from being hired because of the applicant's age. It merely allows the explicit use of the applicant's age in determining whether the individual should be hired. Categorical exclusion on the basis of a characteristic which is a BFOQ is only proper when that characteristic is the sole determinant of expected productivity.

To illustrate this use of the analysis, it is necessary to elaborate upon the example of the employer who is hiring for a job. The employer's decisionmaking model described above is objectively nondiscriminating because neither sex nor age is explicitly included as a determinant of the probability of hiring. This decision-making model and the resulting probability that an applicant is hired are "sex-blind" and "age-blind."

As alternate decisionmaking models consider the following two formulas used to calculate the probability of hiring an applicant:

$$\text{HIRE}^* = (\text{GRAD}/3) + (\text{STAY}/3) + (\text{LIFT}/6) + (\text{SEX}/6)$$

and

$$\text{HIRE}^{**} = (\text{GRAD}/4) + (\text{STAY}/4) + (\text{LIFT}/4) + (\text{SEX}/4)$$

where GRAD, STAY and LIFT are defined as above, and SEX equals 1 if the applicant is male and 0 if the applicant is female. The decisionmaking model used to calculate HIRE* weights the weightlifting qualification only half as much as the model determining the probability HIRE. This fact alone would decrease the sex difference in the probability of being hired. However, the applicant's sex is included as a qualification, weighted to be half as important as educational attainment and willingness to stay on the job for five years. The probability HIRE** is determined by a decisionmaking model which places equal weight on the three job qualifications and the applicant's sex. Using these decisionmaking models, the resulting probability that an applicant is hired is directly determined by the applicant's sex. Both models discriminate against female applicants by lowering their probability of being hired regardless of their qualification for the job (expected productivity).

Since the decisionmaking model used by an employer is rarely as explicitly stated as in the formula used to calculate HIRE, HIRE* or HIRE**, the proof of discrimination is the determination of which model the employer uses in his hiring decisions. Whether the employer uses the nondiscriminatory model which produces the nondiscriminatory hiring probability HIRE or either of the two discriminatory models can be determined by examining the partial correlation between sex and the hiring probability after controlling for the job qualifications. A comparison of the partial correlations resulting from each of the three decisionmaking models is reported in Table III.

TABLE III
RELATIONSHIP BETWEEN SEX AND HIRING PROBABILITY
WITH THREE DECISIONMAKING MODELS

| | Nondiscriminatory Model | Discriminatory Models | |
|--|----------------------------|-----------------------|--------|
| | HIRE | HIRE* | HIRE** |
| Correlation Between Sex and Hiring Probability | .34 | .55 | .74 |
| Partial Correlation Between Sex and Hiring Probability Controlling for Weightlifting Qualification | .00 | .38 | .63 |
| Percent of Sex Difference in Hiring Probability Due to Employer's Discrimination | 0% | 69% | 85% |

As noted above, with the nondiscriminatory probability HIRE, the correlation between sex and the hiring probability is totally a result of the sex difference in the weightlifting job qualification. Thus when the relationship between sex and the hiring probability is controlled for differences in this qualification, the partial correlation of sex and the hiring probability is zero.

The lack of a relationship between sex and the hiring probability, after taking into account the sex difference in the weightlifting qualification, is not evident when either of the two discriminatory decisionmaking models is used by the employer. When the discriminatory hiring probability HIRE* is the applicant's chance of employment, a positive correlation between sex and the hiring probability remains after controlling for differences in qualifications. When the decisionmaking model used by the employer produces the hiring probability HIRE**, the evidence of discrimination is even more pronounced. The partial correlation between sex and hiring probability after controlling for sex differences in job qualifications is 0.63. Eighty-five percent (0.63/0.74) of the sex difference in the probability of being hired is due to factors unrelated to job qualifications, *i.e.*, it is a result of discrimination.

The litigant in an employment discrimination case who can demonstrate that a relationship exists between sex (age, race, ethnicity) and the hiring (wage, promotion, dismissal) rate, after controlling for differences in job qualifications, has demonstrated that the employer is using a discriminatory decisionmaking model.

In practice, the analysis required for this demonstration can be considerably more complex than the example presented here. Productivity is determined by many qualifications, some of which may be intervening variables between productivity and characteristics of sex or age. A further complication is the general lack of measurement of the actual qualifications for a job. Few employers have more than a vague notion of what determines a worker's productivity in a particular job. However, as a result of requirements for employment test validation in the U.G.E.S.P., objective job analyses to determine the qualifications will become more common.³⁸ Once these qualifications are defined and applicants are properly measured to determine whether they possess the qualifications, an objective determination of discrimination as illustrated in the analysis may become a common evidential procedure.

CONCLUSION

The three issues discussed in this paper illustrate the potential use and misuse of quantitative information in employment discrimination cases. It should be evident from these illustrations that litigants in employment discrimination cases need to understand the basic logic of statistical inference, the proper measures to indicate discrimination and the analytic procedures which can be used to demonstrate discrimination.

There is no need for lawyers to become statisticians. The actual analysis and presentation of quantitative information should be left to experts. However, proper evaluation of quantitative information as probative to employment discrimination is an appropriate task for the legal community.

³⁸ See §§ 14B2 & 14D2 of the U.G.E.S.P., 43 Fed. Reg. 38,300, 38,307 (1978).

